Faster, More Accurate Characterization of Proteins and Peptides with Agilent MassHunter BioConfirm Software

Technical Overview

Abstract

Analytical labs need rapid, reliable results to solve challenging problems in recombinant antibody characterization, and confirmation of synthetic peptides. Solving those challenges requires the combination of high performance LC/MS instrumentation with highly optimized software tools, such as the new Agilent MassHunter BioConfirm software version B.03.01.

For intact proteins and synthetic peptides, MassHunter BioConfirm software provides:

- The industry's most accurate protein molecular weights – using Agilent TOF data and maximum entropy deconvolution software
- Unrivaled processing speed for complex intact protein mixtures, via the unique Large Molecular Feature Extraction (LMFE) algorithm that both finds proteins and calculates their molecular weights
- A unique feature-finding and deconvolution algorithm (MFE) that is optimized specifically for the analysis of smaller molecules, such as synthetic peptides and peptides in protein digests

For peptide mapping, MassHunter BioConfirm software includes:

- Sequence editor, which allows easy set-up of sequences and definition of known modifications (global or site-specific, fixed or variable)
- Sequence matcher, which accelerates both the confirmation of the expected protein and the identification of unexpected sequence modifications

Agilent Technologies

Introduction

Characterization of the primary sequence of proteins and synthetic peptides, and intended or unexpected modifications, is critically important in research, process development, and quality control (QC). While traditional protein QC methods such as gel electrophoresis provide only an estimate of molecular weight, mass spectrometry (MS) is a superior analytical technique for confirmation of the molecular weight of both proteins and their variants.

The Agilent 6200 Series Accurate-Mass TOF LC/MS Systems and the Agilent 6500 Series Accurate-Mass Q-TOF LC/MS Systems are able to clearly distinguish between closely related compounds and minimize the chance of false positive results. As more laboratories turn to these accurate and reliable instruments for characterization of biomolecules, they need sophisticated data analysis tools that quickly deliver unambiguous results.

These instruments provide the high resolution and mass accuracy that enable detection of minor changes in protein structure, such as post-translational modifications. They also have a dynamic range of up to five orders of magnitude that permits detection of trace impurities in a recombinant product, and low-level peptides in a protein digest. The instruments can be combined with the Agilent 1290 Infinity LC System for greater LC peak capacity and fastest analyses.

MassHunter BioConfirm software includes a complete set of tools for rapid characterization and confirmation of proteins and synthetic peptides. These tools have been optimized for the Agilent Accurate-Mass TOF and Q-TOF instruments. The software provides several different mass spectral deconvolution algorithms that are optimized for intact proteins and peptides, as well as for different sample complexities (**Figure 1**).

The software's built-in sequence editor/matcher allows creation of peptide and protein sequences (including known modifications). The software uses these sequences to confirm the primary sequence and expected modifications, and to identify unexpected modifications. The tools in MassHunter BioConfirm thus accelerate characterization of recombinant proteins and synthetic peptides, and aid in the identification of unwanted impurities and variants.

Optionally, the Agilent Easy Access software can be combined with MassHunter BioConfirm software for simplified operation in a multi-user, walk-up environment.

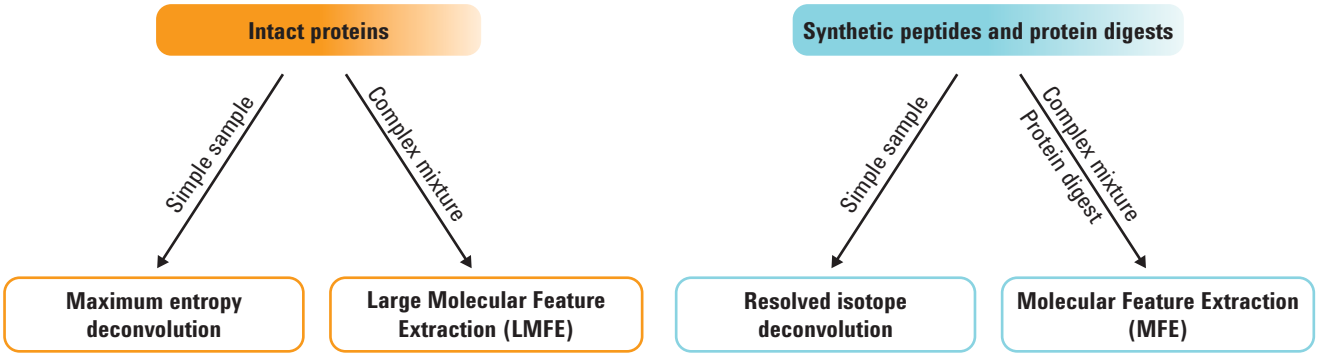


Figure 1. The software provides several powerful deconvolution algorithms that are optimized for different molecular weights and sample complexities.

Tools for intact protein analysis

Accurate molecular weights of proteins

Using the electrospray ionization conditions employed for the analysis of proteins and peptides, mass spectra exhibit a number of multiply-charged peaks (**Figure 2, middle**). Agilent's state-of-the-art deconvolution algorithms convert these complex spectra to simple zero-charge spectra (**Figure 2, bottom**) or masses (**Figure 4**) that directly provide the molecular weights of proteins and their variants.

To deliver the best possible results in the shortest time, MassHunter BioConfirm software includes two easy-to-use complementary deconvolution algorithms for analysis of intact proteins.

Characterization of single proteins and simple protein mixtures

For samples such as recombinant antibodies that consist of a single protein or a simple mixture of proteins, the software provides a maximum entropy deconvolution algorithm to calculate the protein molecular weight(s). As shown in **Figure 2**, this algorithm processes the complex mass spectrum with multiple charge states, and delivers an accurate molecular weight—usually within 25 ppm. If the protein contains impurities or heterogeneities (as was the case in **Figure 3**), the algorithm reveals the molecular weights of the additional components.

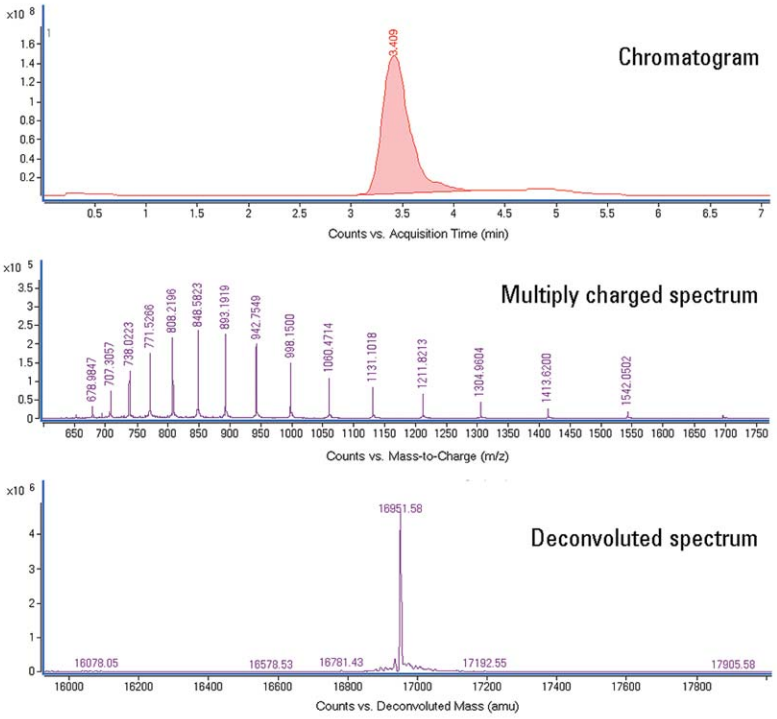


Figure 2. Deconvolution of this myoglobin mass spectrum produces the correct molecular weight at 16,951.58 Da.

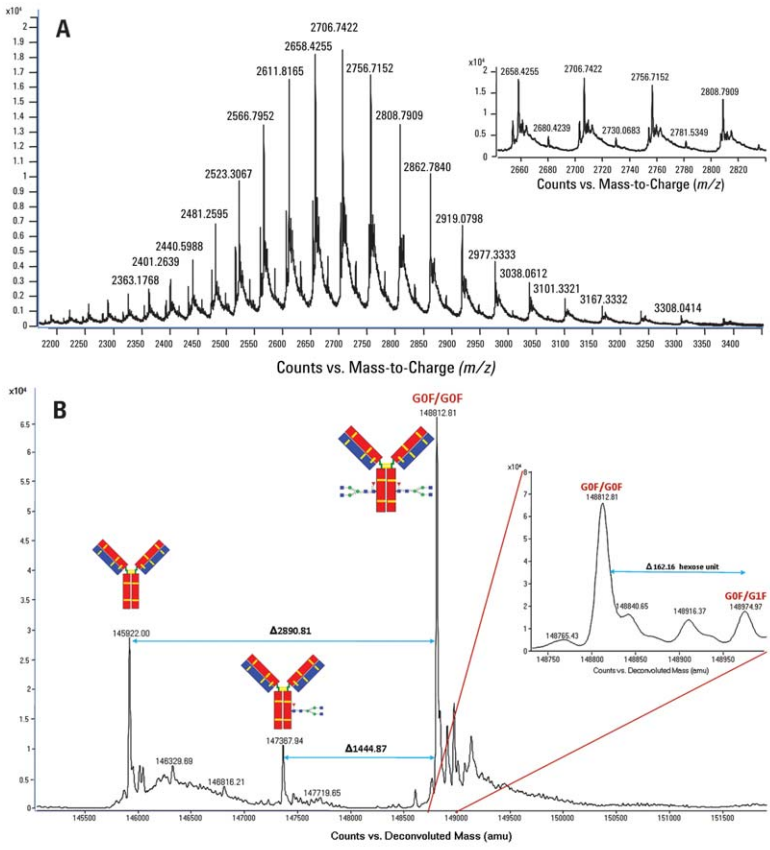


Figure 3. A) Mass spectrum of an intact antibody with inset that shows the expanded view of charge states of the antibody. B) Deconvoluted spectrum of the intact antibody shows multiple forms that contain varying numbers of linked glycans. Inset shows the expanded view of a small amount of G1F glycan moiety (addition of hexose unit to G0F glycan moiety) in the monoclonal antibody. For details, see Agilent publication number 5990-3445EN.

Characterization of proteins in complex mixtures

For samples that contain a large number of proteins, the most efficient method of data analysis uses MassHunter's unique Large Molecule Feature Extraction (LMFE) algorithm to find all the proteins and deconvolute their spectra. This software is much faster than maximum entropy deconvolution when samples contain hundreds of proteins. This innovative algorithm has its origins in the Molecular Feature Extraction (MFE) algorithm for small molecules (see sidebar). Ions that correspond to different charge states of the same protein elute at the same time in the LC/MS chromatogram. The LMFE algorithm recognizes the relationship between those ions, groups them into a single protein compound, and calculates the protein molecular weight. **Figure 4** shows an example of the results from LMFE.

The advantage of using LMFE versus the maximum entropy deconvolution algorithm is that LMFE can be applied to the entire LC/MS chromatogram without prior selection of the mass spectra to be deconvolved. When using maximum entropy deconvolution, the user must create average mass spectra across various chromatographic peaks or time ranges, execute the algorithm on each of the spectra, and specify the expected mass ranges.

Molecular Feature Extraction

The Molecular Feature Extraction (MFE) algorithm is a powerful compound-finding algorithm that locates individual compounds (molecular features) in complex samples, even when compounds are not mass resolved. Rather than designate compounds based exclusively on two-dimensional chromatographic peak information, MFE locates ions that are covariant (rise and fall together in abundance) in the three-dimensional LC/MS data set and that are logically related by charge-state envelope, isotopic distribution, and/or the presence of adducts and dimers. It assigns all related ions to a single compound. Using this approach, the MFE algorithm can distinguish multiple co-eluting compounds that would appear as a single peak in a total ion chromatogram.

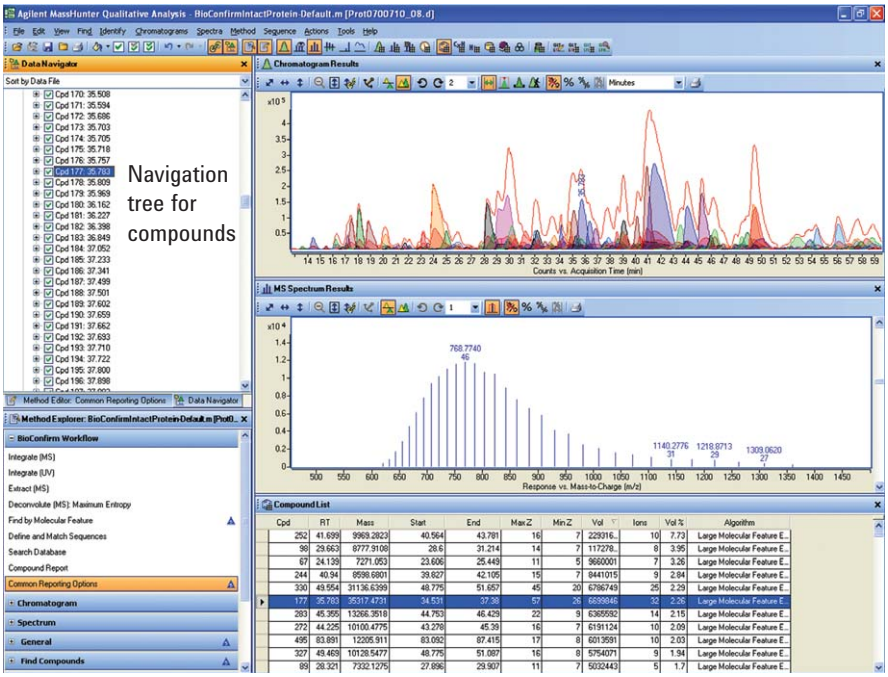

Navigation tree for compounds
Extracted compound chromatograms (ECCs) for proteins
Spectrum for protein compound
Protein masses in compound list

Figure 4. The LMFE algorithm automatically finds all the protein compounds in a complex LC/MS analysis and lists their masses—saving hours of manual data analysis.

The superior performance of LMFE for rapid location and deconvolution of all the proteins in a complex LC/MS sample can be demonstrated by the 140 min run of intact proteins from *E. coli* (**Figure 5**). **Table 1** compares the performance of LMFE versus maximum entropy deconvolution, and shows the superiority of LMFE when the sample contains hundreds of proteins. The LMFE algorithm is approximately six times faster than maximum entropy deconvolution in this example. It also finds proteins that co-elute, which results in more comprehensive sample characterization.

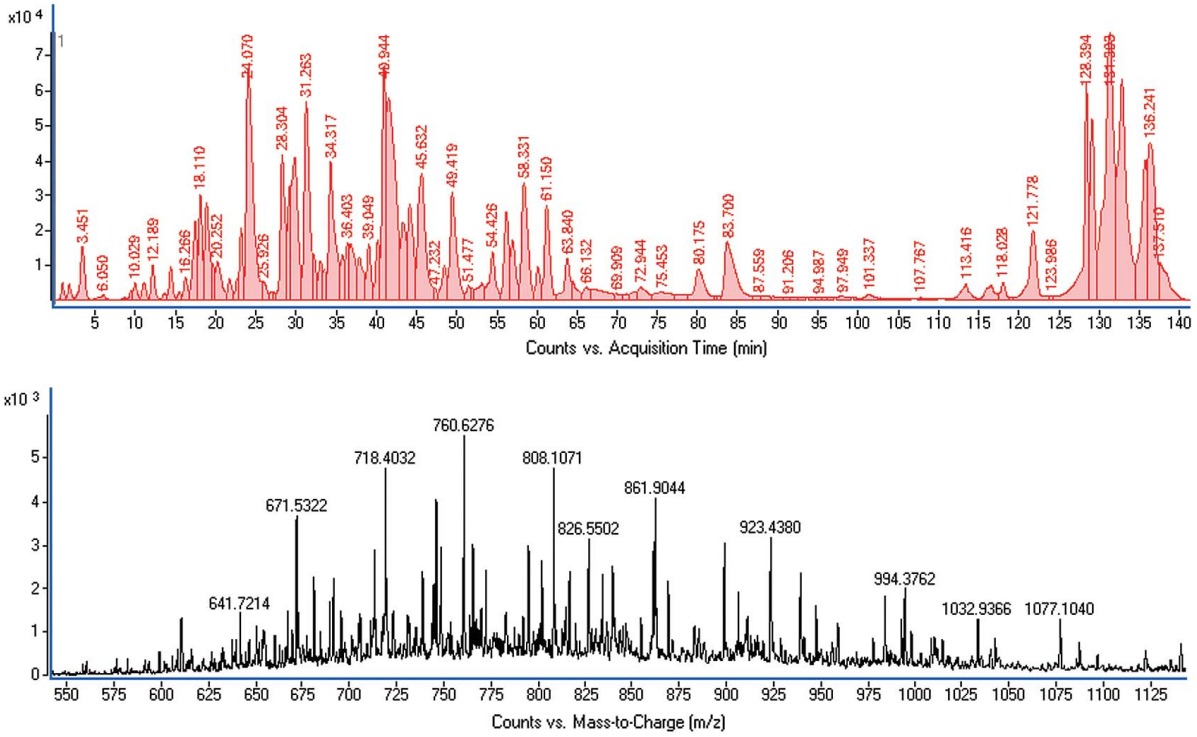


Figure 5. For this complex mixture of E. coli proteins, the LMFE algorithm identified three times the number of proteins and took one-sixth the time when compared with the maximum entropy deconvolution.

Algorithm	Total time	Number of proteins found
Maximum entropy	90 min (to integrate and extract 73 peak spectra, then deconvolute each spectrum)	192
LMFE	15 min	597

Table 1. Comparison of maximum entropy deconvolution versus LMFE for the processing of an LC/MS analysis of E. coli proteins.

Tools for synthetic peptide analysis and peptide mapping

Resolved isotope deconvolution for determining accurate molecular weights of peptides

With peptides that have molecular weights less than 4 kDa, MassHunter BioConfirm uses an optimized algorithm for resolved isotope deconvolution (**Figure 6**). This algorithm produces zero-charge centroided spectra from MS and MS/MS data of compounds with resolved isotope distributions, which allow deduction of the charge state directly from the *m/z* difference between the isotopes. Agilent TOF and Q-TOF instruments with 4 GHz data acquisition can resolve isotopes for compounds with molecular weights up to about 8 to 20 kDa.

Rapid analysis of complex peptide mixtures via MFE

For peptide mixtures that contain many components, manual analysis of LC/MS results can become time consuming. To automate the process, MassHunter Qualitative Analysis provides the Molecular Feature Extraction (MFE) algorithm (see sidebar). In a one-click operation, it both finds the peptides in the LC/MS chromatogram, and determines their masses via resolved isotope deconvolution. MFE not only uses the resolved isotopes to directly determine the charge states from the peptides, but then also looks for related ion clusters that represent the same peptide with neighboring charge states. MFE is the fastest way to get results for complex mixtures, saving hours of labor relative to competitive software applications.

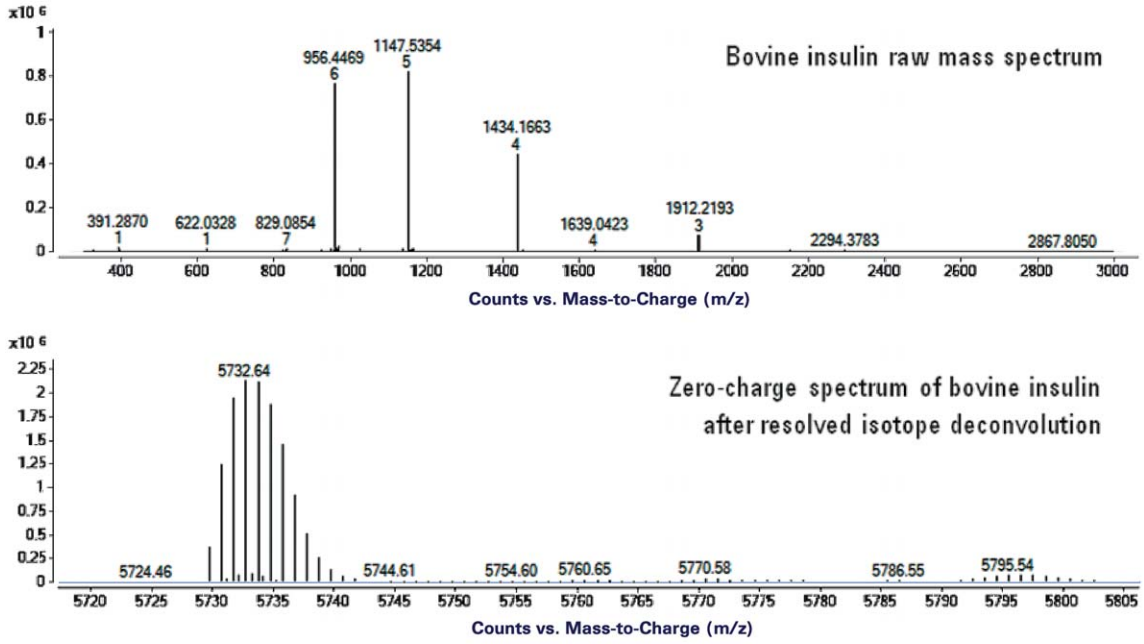


Figure 6. Resolved isotope deconvolution produces a monoisotopic mass of 5729.64 for bovine insulin, in close agreement with the calculated monoisotopic mass of 5729.60.

Peptide mapping: time saving tools match sequences to compounds

To confirm and characterize a protein at the level of digested peptides, the software uses an approach known as "peptide mapping." The software first calculates the masses of peptides from a theoretical digestion of a known amino acid sequence. It then compares these theoretical mass values with the masses detected in the LC/MS analysis of the peptide digest. The sequence editor/matcher within MassHunter BioConfirm aids with both the protein confirmation and the identification of impurities and modifications.

To carry out peptide mapping, the user first creates or imports the sequence into the editor/matcher and then specifies the protease used to digest the protein. The software performs a theoretical digestion. After the LC/MS analysis of the digest, the software matches the predicted masses from the sequence against the deconvoluted peptide masses in the compound list generated by the Molecular Feature Extraction (MFE) algorithm.

Figures 7 and **8** show results of peptide mapping for a digest of an Immunoglobulin G (IgG) on the Agilent 1290 Infinity LC and the Agilent 6540 Accurate-Mass Q-TOF LC/MS. In this example, the LC/MS analysis took less than three minutes, and gave complete (100%) amino acid sequence coverage leading to amino acid confirmation of the protein.

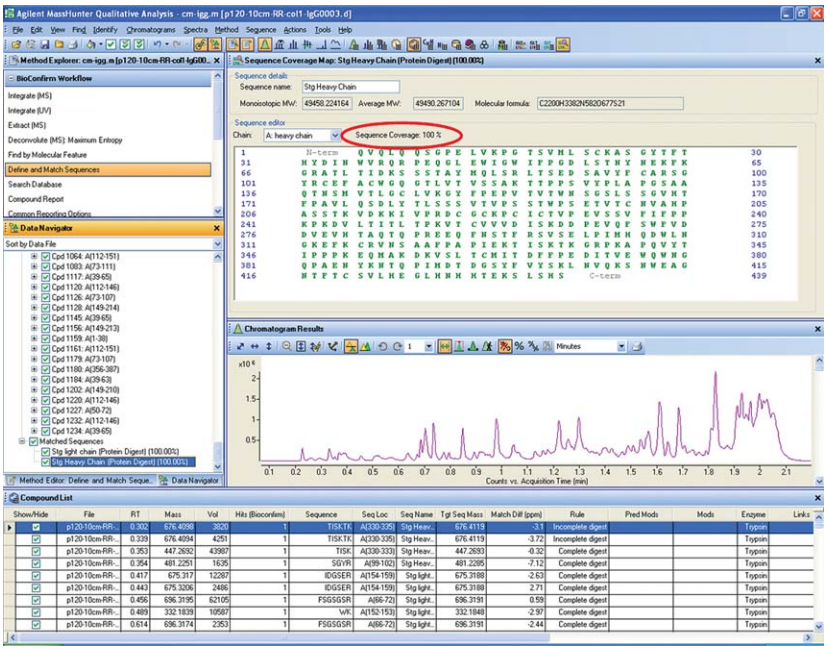


Figure 7. For peptide mapping, the software calculates coverage and highlights the portions of the sequence that match the sample. In this case, the Agilent system delivered 100% coverage with only a three-minute LC/MS run.

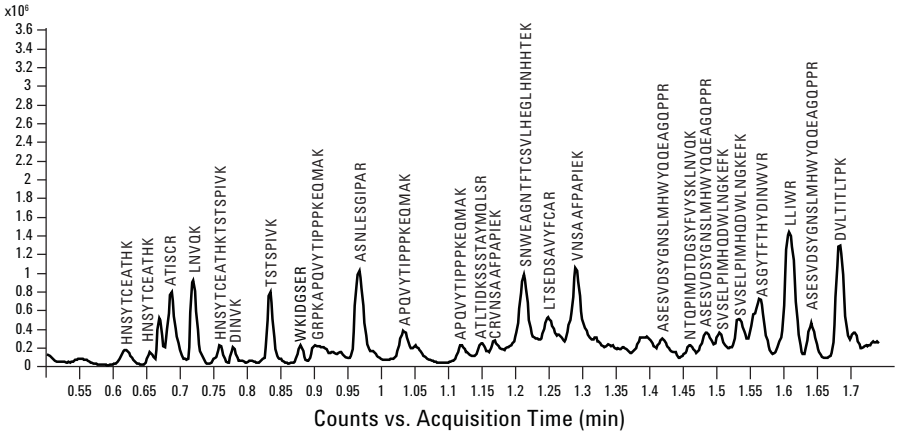
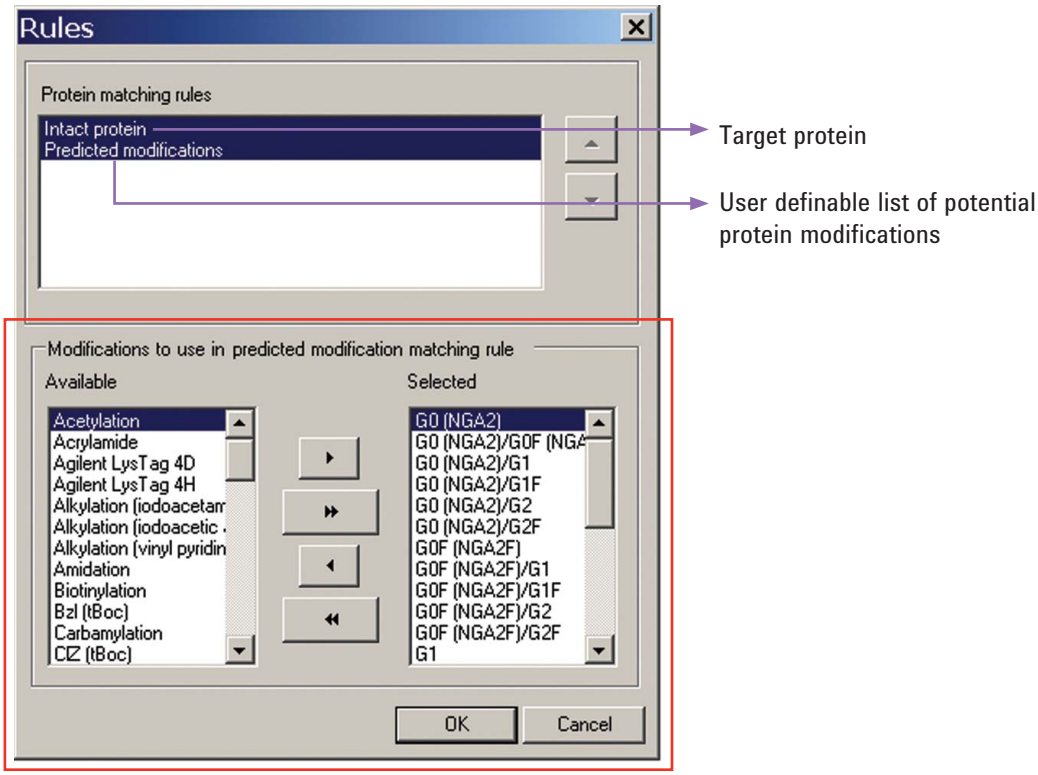


Figure 8. The software automatically labels the MS chromatogram with the amino acid sequences of the peptides.

Troubleshoot problems and characterize peptide and protein modifications

In process development or quality control, MassHunter BioConfirm software helps identify failed syntheses and those with insufficient purity. The sequence editor/matcher shows portions of the sequence where no matches were found in the LC/MS analysis. These are potential locations of incorrect sequences, mutations, or post-translational modifications. The sequence editor/matcher allows the investigator to specify potential modifications to a protein or synthetic peptide and then searches for evidence for those modifications in the LC/MS results. For digests, the user can specify whether missed cleavages or point mutations are allowed, and can select the post-translational modifications that should be searched.

The sequence editor/matcher allows one to specify whether a modification should be searched for each applicable amino acid, or assumed to be site-specific. It also allows the user to specify whether a modification can be variable, so that it is not present in every peptide or protein molecule in the sample. **Figure 9** shows an example of how the software can automatically detect glycosylated variants of the main protein at the intact protein level.



Mass	Vol	Hits (Bioconfirm)	Sequence	Seq Loc	Seq Name	Tgt Seq Mass	Match Diff (ppm)	Rule	Pred Mods
148217.97...		2	DIQMTQSPSSL...	A(1-215)...	IgG	148219.0064	-6.95	Protein modifica...	2*G1
148381.008		1	DIQMTQSPSSL...	A(1-215)...	IgG	148381.1496	-0.95	Protein modifica...	1*G1/G2
148057.92...		1	DIQMTQSPSSL...	A(1-215)...	IgG	148056.8631	7.16	Protein modifica...	1*G0 (NGA2)/G1
148542.669		1	DIQMTQSPSSL...	A(1-215)...	IgG	148543.2929	-4.2	Protein modifica...	2*G2 (NA2)

Figure 9. The software can automatically find standard glycoforms, so the analyst does not need to take the time to add the glycan modifications to the protein sequence.

Additional functionality to enhance productivity

Peptide AM(RT) database matching for automated QA/QC

The MassHunter Qualitative Analysis software enables matching of peptides to a user created accurate mass (AM) database with optional retention time (RT). The AM(RT) database can be in comma-separated value (.csv) format. It contains the amino acid sequence, the theoretical accurate mass, and optionally the retention time for each of the peptides. When using a standardized chromatography method, addition of retention time allows one to unambiguously distinguish peptides with the same molecular formula and amino acid content, but different amino acid sequences.

Peptide database matching is an excellent QA/QC tool for repetitive analyses of the same protein, such as digests of recombinant proteins. The investigator can use automated Molecular Feature Extraction to quickly find all the peptides in an LC/MS analysis of a protein digest. Then the accurate masses in the compound list from MFE can be searched against the custom database to find matching peptide sequences. **Figure 10** shows an example of the results of this completely automated analysis.

After the search is complete, the software replaces the compound names in the Data Navigator with the peptide sequences. The Compound List in the lower right of Figure 10 shows the matched peptides, retention times, and mass differences in ppm. The ability to create and search reference databases that contain accurate masses of peptides saves many hours of time for routine confirmation.

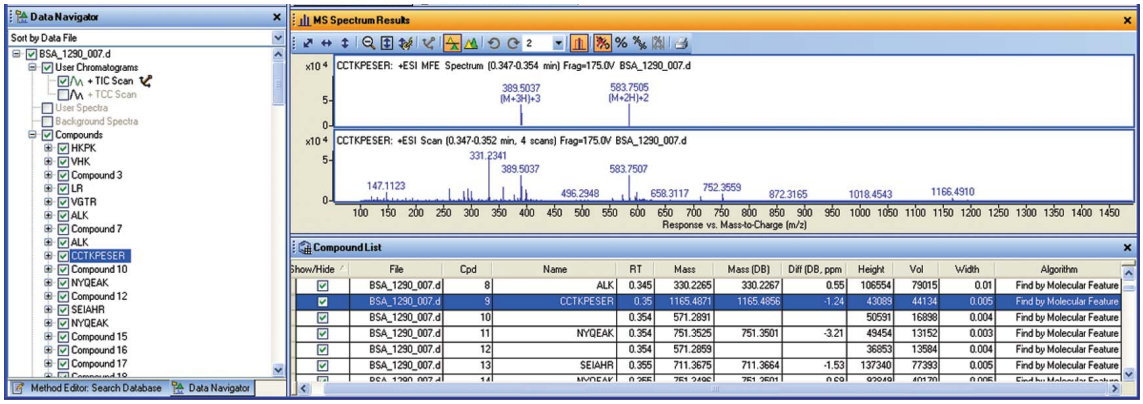


Figure 10. The MassHunter Qualitative Analysis software allows creation of a custom peptide database, which can be searched for matches to masses of peptides found in the LC/MS results.

Compound-based navigation for fastest review of results

Because MassHunter BioConfirm is an add-in for the Agilent MassHunter Qualitative Analysis software, it inherits a unique capability for compound based navigation that makes it much quicker to review results for proteins and peptides (Figure 4). For example, when LMFE finds a protein, it lists the protein as a compound in the navigation tree in the Data Navigator pane. Under each compound node, the software provides one-click navigation to the:

- Centroided ion set that provides evidence for the protein
- Extracted compound chromatogram (ECC) created from all the ions in that set, over the elution range of that compound
- Deconvoluted spectrum
- Row in the Compound List that gives all processing results for the specific compound (see bottom of Figure 4)

Easy Access software for simple, automated walk-up operation

Agilent Easy Access software provides a simplified user interface for submission of samples for protein and peptide confirmation and characterization. Users need only to log in, enter sample information, and choose from a list of preconfigured methods for both acquisition and data processing. The software tells them where to place the samples and approximately when the analysis will be complete, and provides detailed confirmation reports by hard copy, web page, or e-mail.

Easy Access software can use preconfigured data acquisition and analysis methods that Agilent supplies, or custom methods created by experienced users. With MassHunter Easy Access and BioConfirm software, less experienced users can take advantage of the power of TOF-MS for routine analyses of recombinant proteins and synthetic peptides.

Conclusion

- Fully automated and comprehensive data processing
  Agilent TOF and Q-TOF systems provide an exceptional combination of mass accuracy, resolution, sensitivity, and dynamic range. MassHunter BioConfirm software complements the high-quality data produced by those instruments, and allows fully automated and comprehensive data processing specifically for this application. The software supports workflows to confirm and characterize proteins both at the intact and digest level.
- Peptide Mapping
  MassHunter BioConfirm allows the setup of sequences for one or multiple proteins or synthetic peptides and the definition of modifications (global or site-specific). It also allows extensive peptide mapping capabilities including visualization of matched sequence coverage, including modifications.
- Deconvolution software includes powerful deconvolution algorithms optimized to determine molecular weights from recombinant proteins and synthetic peptides, in both simple and complex samples
  1. A maximum entropy deconvolution algorithm is available for intact proteins and allows excellent mass accuracy.
  2. MassHunter's unique LMFE algorithm automatically and rapidly finds intact proteins and their variants in complex samples, and deconvolutes their spectra.
  3. A resolved isotope deconvolution algorithm determines the molecular weights of peptides.
- Highest productivity
  1. MFE and LMFE rapidly simplify peptide and protein spectra
  2. Fast compound based navigation via dynamic link of compound mass spectra, chromatograms, and results
  3. Indentification of peptides in QA/QC via AM(RT) database search (optional)
  4. Easy Access software for easy-to-use, fully automated, walk-up confirmation of peptide and protein molecular weights.

The combination of these Agilent products significantly increases productivity for busy scientists in the biopharmaceutical industry who are pressed to characterize and confirm more proteins and peptides in less time and with greater confidence.

www.agilent.com/chem/masshunter

This item is intended for Research Use Only. Not for use in diagnostic procedures.

Agilent Technologies shall not be liable for errors contained herein or for incidental or consequential damages in connection with the furnishing, performance or use of this material.

© Agilent Technologies, Inc. 2010
Printed in the U.S.A. April 27, 2010
5990-5096EN

Agilent Technologies